Statistical modelling of air pollution and the challenges of communication

Carl Scarrott, University of Canterbury, New Zealand

CASI2018

Statistical modelling of air pollution and the challenges of communication

Carl Scarrott, University of Canterbury, New Zealand

CASI2018

Statistical modelling of air pollution and the challenges of communication

Carl Scarrott, University of Canterbury, New Zealand

CASI2018

Statistical modelling of air pollution and the challenges of communication

Carl Scarrott, University of Canterbury, New Zealand

CASI2018

Statistical modelling of air pollution and the challenges of communication

Carl Scarrott, University of Canterbury, New Zealand

CASI2018

Context

- Winter pollution - particulate matter $\leq$ 10 microns (PM_{10})
- Predominantly home heating (e.g. open fires, inefficient log burners)
- PM_{10} concentrations well above WHO guidelines (e.g. daily average of 50 $\mu g/m^3$)

- Strongly linked to various medical conditions, incl. respiratory and cardiac problems
- Linked to increased mortality
- Social costs estimated at over NZ$4 billion per year (€2.5 billion)

National Environmental Standards

- Govt set National Environmental Standards for Air Quality:
  - **Daily average** PM_{10} concentrations (midnight to midnight)
  - maximum number of exceedances of target level $50\mu g/m^3$
    - no more than 1 exceedance by 2020
- Regional councils responsible for implementation
- Various central, regional and local government initiatives to achieve these targets

- Winter PM_{10} concentrations are mainly driven by:
  - emissions - meteorology and number of burners used
  - meteorology directly

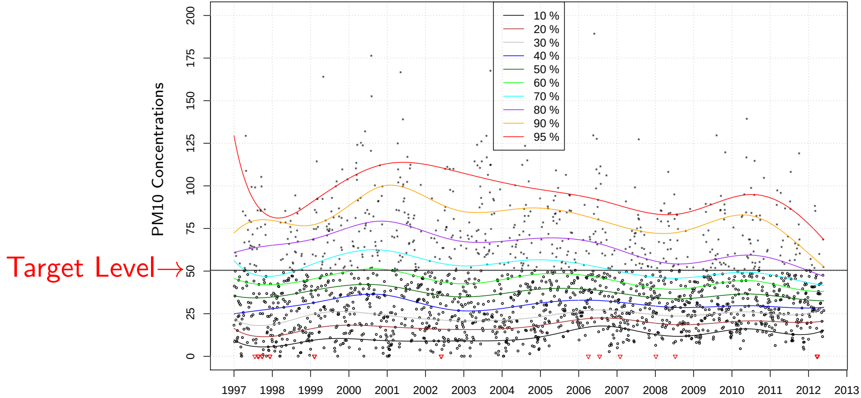
Key Questions of Interest

1. What is the current situation and have the govt initiatives been effective?
2. What % reduction in the emissions (e.g. inefficient log-burners) is needed to achieve the targets?

In more statistical terminology:

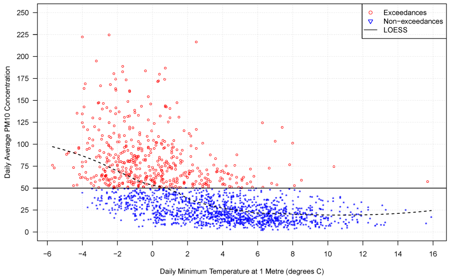
1. Statistical model of the meteorological impacts on the concentrations along with historical (slowly varying) trend to capture changes in extent of burner usage.
   - Is their evidence for a decreasing trend since 2003?
2. Estimate % reduction in the concentrations needed to reach future targets with a certain probability, taking into account random meteorological variation and model uncertainty

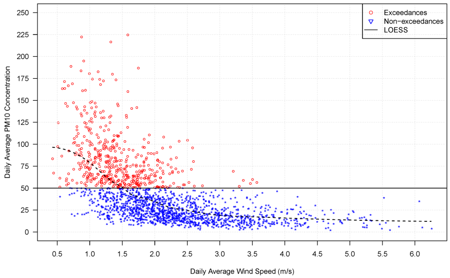
Daily Average Winter PM_{10} Concentrations



- Typical exceedance night: cold, calm and no rain (possibly with temperature inversion)
- Only winter data (May-August) as few exceedances in summer and these are caused by different factors
- Notice spurious lack of low concentrations since 2005

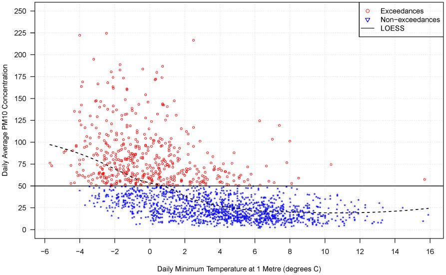
Meteorological Impacts

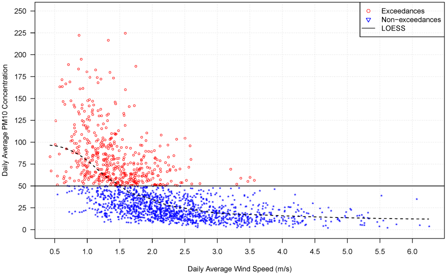




- Key predictors: temperature and wind speed
- Negative exponential form so model mean on log-link scale
- Secondary predictors: rainfall, soil temperature, wind direction, relative humidity and cloud cover/height

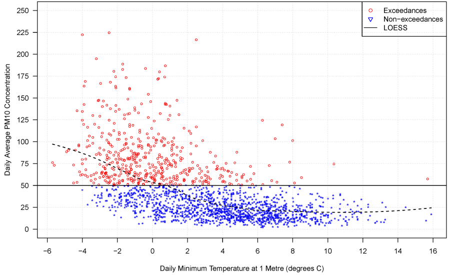
Meteorological Impacts

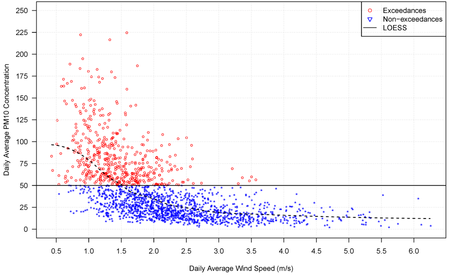




- Hourly meteorological data
- Aggregated to daily statistics: mean, **minimum**, **maximum** and **standard deviation**

Meteorological Impacts





- Extra complication:
  - Legal definition 24 hour is midnight to midnight
  - Physical reality ∼3PM to ∼3PM
- Need previous evening's meteorology

GAMM for Daily PM_{10} Concentrations

- Generalized additive mixed model on log-link scale $g(\mu) = \log(\mu) = \eta$

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(t) + f_2(t) + \ldots + f_k(t)$$

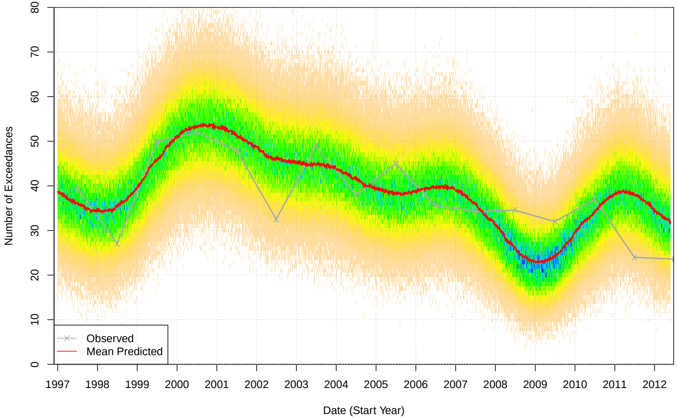
- Meteorological fixed effects as linear predictors (following some careful model selection)
- Slowly varying function of time t:
  - thin plate regression splines ($\sim$ 7-8 d.f.)
- Distribution family:
  - Normal(μ, σ^2) - constant variance? Negative concentrations?
  - Gamma(α, β) - variance $\propto$ mean and positive concentrations (physically sensible but not robust to data issues)
- AR(1) to capture residual autocorrelation

Simulations to Visualise Model Fit

- Take estimated smooth trend on a particular day and fix it for whole winter period
- Simulate potential winter weather patterns using historical meteorological data vectors:
  - Randomized block length (stationary) bootstrap
  - Average block length of 5 days
  - Gives simulated weather for all 123 days of winter
  - Repeat 10,000 times
- Apply GAMM to get predicted concentrations $\hat{\mu}_{sim} = \exp(\hat{\eta})$
- Simulate "observable values" of PM_{10} concentrations from GAMM distribution (normal, gamma, ...)
- For each "simulated winter" extract relevant statistics:
  - number of exceedances
  - second largest PM_{10} concentration
  - fourth largest PM_{10} concentration

Results for Number of Exceedances

- Simulated statistics were much easier to explain to scientists and policymakers:



- Satisfactory model diagnostics and estimated trend is consistent with observed behaviour
- Large deviation in 2012 due to partially observed year

Simulate From Model For Forecasting

- Back to second question...
  - Estimate % reduction in the concentrations needed to reach future targets with a prescribed probability, taking into account random meteorological variation and model uncertainty

- Use simulations to estimate total % reduction required
- Prescribed probabilities of 50% and 95%
- Use estimated trend in current year as baseline level
- Inverse problem to find % reduction needed to get 50% of "simulated winters" to have at most 1 exceedance by 2020

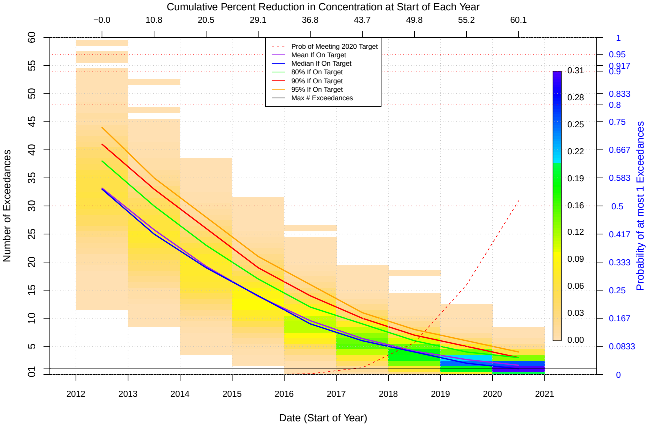
Simulate From Model For Path to Compliance!

- Using the total % reduction needed by 2020, consider potential "**paths to compliance**"
  (sorry for the management speak!)
- Apply linear trend on log-scale, to give **constant % reduction each year**
- Use simulated distribution of:
  - number of exceedances
  - second largest concentrations

  to get tolerance limits per year when "city is on target"
- i.e. to allow for fact that city could be on target, but may have a particularly cold winter by chance

Path to Compliance With 50% Probability

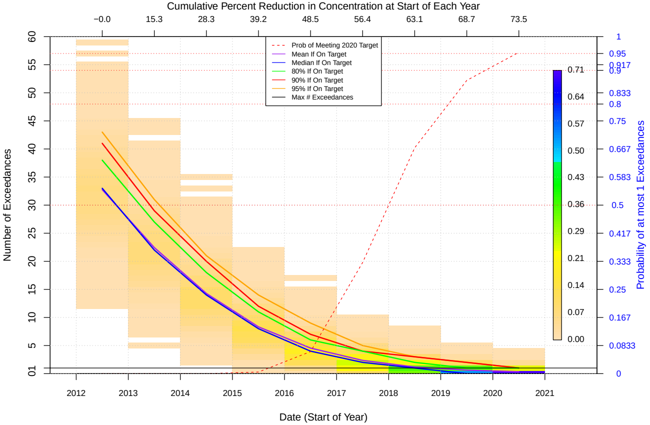
- Total estimated reduction of 60.1% of PM_{10} concentrations
- Annualised value of 10.8%



- Aim at (median, blue line) at 1 exceedance by 2020, as expected
- Green/red/orange lines are tolerable limits on number of exceedances if on target, to allow for worst 20%, 10% or 5%

Path to Compliance With 95% Probability

- Total estimated reduction of 73.5% of PM_{10} concentrations
- Annualised value of 15.3%



- Aim at less than 1 exceedance by 2020!
- Second largest PM_{10} concentration aim of $42\mu g/m^3$ for 95% chance of meeting 2020 target

Summary for Complying by 2020

		2013	2014	2015	2016	2017	2018	2019	2020
50% Chance of Meeting 2020 Target									
Annual % Reduction		**10.8**							
Cumulative % Reduction		10.8	20.5	29.1	36.8	43.7	49.8	55.2	**60.1**
Target Values	# Exceed.	26	19	14	9	6	4	2	1
	Second Larg.	88	80	73	67	62	58	54	50.5
90% Up Limit	# Exceed.	33	25	19	14	10	7	5	3
	Second Larg.	99	90	83	76	70	66	61	58
95% Chance of Meeting 2020 Target									
Annual % Reduction		**15.3**							
Cumulative % Reduction		15.3	28.3	39.2	48.5	56.4	63.1	66.7	**73.5**
Target Values	# Exceed.	23	14	8	4	2	1	0	0
	Second Larg.	84	74	65	59	53	49	45	42
90% Up Limit	# Exceed.	29	20	13	8	4	3	2	1
	Second Larg.	95	84	74	66	60	55	52	48

Summary and Further Work

- Interesting application bringing together statistical modelling and simulation
- Provides scientific justification to inform government policy

- Earthquakes changed priorities
- From air pollution viewpoint, major improvement to housing/heating in Christchurch due to rebuild

- Further work to do:
  - uncertainty due to "estimating current state"
  - robustness to data issues
  - alternative modelling approaches (exceedances are extremes)
  - multi-site extensions